

Overcoming the false-minima problem in direct methods: Structure determination of the packaging enzyme P4 from bacteriophage ϕ 13

Article (Published Version)

Meier, C, Mancini, E J, Bamford, D H, Walsh, M A, Stuart, D I and Grimes, J M (2005) Overcoming the false-minima problem in direct methods: Structure determination of the packaging enzyme P4 from bacteriophage ϕ 13. *Acta Crystallographica Section D: Biological Crystallography*, 61 (9). pp. 1238-1244. ISSN 0907-4449

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/52561/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Overcoming the false-minima problem in direct methods: structure determination of the packaging enzyme P4 from bacteriophage ϕ 13

Christoph Meier,^a Erika J. Mancini,^a Dennis H. Bamford,^b Martin A. Walsh,^c David I. Stuart^a and Jonathan M. Grimes^{a*}

^aDivision of Structural Biology, The Henry Wellcome Building for Genomic Medicine, Oxford University, Roosevelt Drive, Oxford OX3 7BN, England, ^bInstitute of Biotechnology and Department of Biological and Environmental Sciences, Vikki Biocenter, PO Box 56 (Viikinkaari 5), 00014 University of Helsinki, Finland, and ^cBM14, c/o ESRF, 6 Rue Jules Horowitz, BP 220, 38043 Grenoble CEDEX 9, France

Correspondence e-mail:
jonathan@strubi.ox.ac.uk

Received 6 April 2005
Accepted 22 June 2005

The problems encountered during the phasing and structure determination of the packaging enzyme P4 from bacteriophage ϕ 13 using the anomalous signal from selenium in a single-wavelength anomalous dispersion experiment (SAD) are described. The oligomeric state of P4 in the virus is a hexamer (with sixfold rotational symmetry) and it crystallizes in space group *C*2, with four hexamers in the crystallographic asymmetric unit. Current state-of-the-art *ab initio* phasing software yielded solutions consisting of 96 atoms arranged as sixfold symmetric clusters of Se atoms. However, although these solutions showed high correlation coefficients indicative that the substructure had been solved, the resulting phases produced uninterpretable electron-density maps. Only after further analysis were correct solutions found (also of 96 atoms), leading to the eventual identification of the positions of 120 Se atoms. Here, it is demonstrated how the difficulties in finding a correct phase solution arise from an intricate false-minima problem.

1. Introduction

In a crystallographic diffraction experiment the intensities of the diffracted X-rays are measured, whereas the associated relative phases are lost. This is referred to as 'the phase problem' as without the relative phase information a map of the electron density which gives rise to the diffraction pattern cannot be calculated. Classically, the phase problem has been overcome by using the method of multiple isomorphous replacement (MIR; Green *et al.*, 1954). Here, heavy atoms are soaked into the protein crystal which, owing to their mass, introduce measurable intensity differences. These differences can be used to deduce the heavy-atom positions and phases can then be calculated. A major drawback of MIR is that the search for suitable heavy-atom derivatives often fails. The technique has for the most part been superseded by the multiple-wavelength anomalous diffraction (MAD) method, which relies on only one crystal containing a heavy atom that provides a suitable source of anomalous diffraction. The diffraction experiment then consists of collecting multiple data sets, usually two or three, around the X-ray absorption edge of the heavy atom. The phase information is obtained from intensity differences arising from the energy-dependent variation in scattering from the anomalously scattering heavy atoms. Incorporation of selenium in the form of selenomethionine as a source of measurable anomalous scattering, together with the general uptake of cryo-crystallography and the relative ease of access to MAD beamlines at modern

synchrotrons, has catapulted the MAD technique to the method of choice for solving protein structures over the last decade (Hendrickson, 1991). Experiments that use data from a single wavelength (single-wavelength anomalous dispersion; SAD) have more recently become popular as it has become clear that a single-wavelength experiment is in most cases sufficient to produce an interpretable electron-density map. In both MIR and SAD/MAD experiments, it is necessary to locate the positions of the subset of 'heavier' atoms so that reference phases can be derived. In the case of selenomethionine experiments it is not uncommon that there are more than ten methionines in a protein subunit of modest size, so that in the presence of non-crystallographic symmetry (NCS) several dozen seleniums often have to be found. Patterson functions have traditionally been used to solve the positions of the heavy atoms, but beyond about 20 or 30 sites automated Patterson interpretation becomes increasingly difficult (the number of peaks increases as the square of the number of atoms). Here, direct methods have come to the rescue since optimized algorithms have been developed that routinely solve the structure of hundreds of atoms; indeed, lysozyme, with 1001 non-H atoms, has been solved *ab initio* using these approaches (Deacon *et al.*, 1998). Direct methods use probabilistic phase relationships and the assumption of equal and resolved atoms to derive reflection phases from a single set of measured intensities. The introduction of dual-space iteration techniques (Miller *et al.*, 1993), commonly known as 'Shake and Bake', has significantly extended the complexity of the structures that can be solved, essentially by adding an atomic peak-picking procedure in real space to reciprocal-space phase refinement. This constraint in real space helps to prevent the propagation of errors or overly consistent phase sets and is essentially density modification incorporating atomicity (Sheldrick *et al.*, 2001). These algorithms are implemented in the computer programs *SnB* (Weeks & Miller, 1999) and *SHELXD* (Sheldrick & Schneider, 1997), which are considered 'state of the art' for solving complex problems by direct methods (Sheldrick *et al.*, 2001).

The implementation of the *Shake-and-Bake* algorithm is similar in both *SnB* and *SHELXD* and involves cyclical iteration between reciprocal and real space. From the largest ~15% of normalized amplitudes (*E* values), a set of all possible structure-factor triple invariants is derived. These invariants are chosen so that the sum of their indices is zero and when the *E* values are all large the most probable value of the sum of the triplet phase is zero. Random trial structures are used to generate initial test phases and these phases are refined using the phase probability constraint in *SnB* by minimizing the minimum function and in *SHELXD* by phase expansion from the most reliable phases using the tangent formula. The procedure then switches into real space to impose the constraint of *N* atomic sites of approximately equal scattering power. This enables chemically sensible criteria such as minimal atomic distances and the constraint of an approximate number of sites to be applied. The process iterates for a number of cycles. Solutions are recognized in *SnB* by

a drop in the minimum function R_{\min} , reflected in a bimodal distribution of correct and incorrect solutions, and in *SHELXD* by a clear increase in the correlation coefficient (CC). In the experience of the authors of *SHELXD* when there are solutions with CCs greater than 30% that are separate from the other solutions, the substructure has been solved (T. R. Schneider and G. M. Sheldrick, personal communication).

Despite the limited resolution and noise within the data, this dual-space direct-method approach works well for SAD/MAD experiments, because the problem of describing the substructure of anomalous scatters is still highly redundant (the distances between atoms of the substructure being considerably greater than the resolution of the data). Direct methods are very sensitive to data quality and so it is generally recommended that highly redundant data is collected (Schneider & Sheldrick, 2002). The power of these approaches means that it is now possible, with accurately measured data, to solve substructures of well over 100 atoms (von Delft *et al.*, 2003). Nevertheless, direct-method algorithms can yield incorrect phase solutions with crystallographic quality indicators that resemble correct solutions (so-called false minima). These are especially common in space group *P1*, where the incorrect solutions typically have all phase angles calculated to be zero, and occur less frequently in other space groups, except for complex systems with a large number of atoms (Xu *et al.*, 2000). Here, we describe an example of the false-minima problem which arose from the presence of more than one copy of a symmetric oligomer within the crystallographic asymmetric unit. Although aspects of this case are unusual, we believe that as macromolecular crystallography tackles more complex problems, detecting and correcting such behaviour will become important.

2. The P4 packaging NTPase from the *Cystoviridae* family of bacteriophages

The *Cystoviridae* are a family of enveloped double-stranded RNA (dsRNA) bacteriophages which includes the viruses $\phi 6$ through to $\phi 13$. One of the key steps in the assembly of these bacteriophages is the packaging of the viral genome inside the nascent procapsid. Genetic and biochemical studies have shown that this process is mediated by the viral protein P4, a non-specific ATPase located on the surface of the procapsid (for a review, see Mindich, 1999). To shed light on RNA packaging during the assembly process, we are studying the structure of the P4 protein from different members of the *Cystoviridae* using X-ray crystallography. We have crystallized several P4s and reported the structure of one from phage $\phi 12$ (Mancini *et al.*, 2003; Mancini, Kainov, Grimes *et al.*, 2004; Mancini, Kainov, Wei *et al.*, 2004). These proteins share little sequence similarity and although they are all apparently hexameric, structure determination by molecular replacement using a P4 from a related virus is not generally feasible. Therefore, we embarked on phase determination of P4 from $\phi 13$ using SeMet labelling and MAD/SAD phase analysis.

Table 1

Crystallographic data-collection and processing statistics.

Values in parentheses refer to the highest resolution shell.

X-ray source	ESRF BM14
Wavelength (Å)	0.9791
Resolution range (Å)	20.0–2.5 (2.6–2.5)
Space group	C2
Unit-cell parameters (Å, °)	$a = 274.1$, $b = 260.7$, $c = 137.2$, $\beta = 90.2$
Mosaic spread (°)	0.4
Observations	3087364
Unique reflections	324031
Completeness (%)	99.1 (93.7)
$I/\sigma(I)$	9.9 (1.4)
$R_{\text{merge}}^{\dagger}$ (%)	10.2 (72.3)
No. of hexamers in AU	4

$\dagger R_{\text{merge}} = \sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)$, where $I_i(h)$ is the i th measurement and $\langle I(h) \rangle$ is the weighted mean of all measurements of $I_i(h)$.

3. Crystallization and X-ray analysis

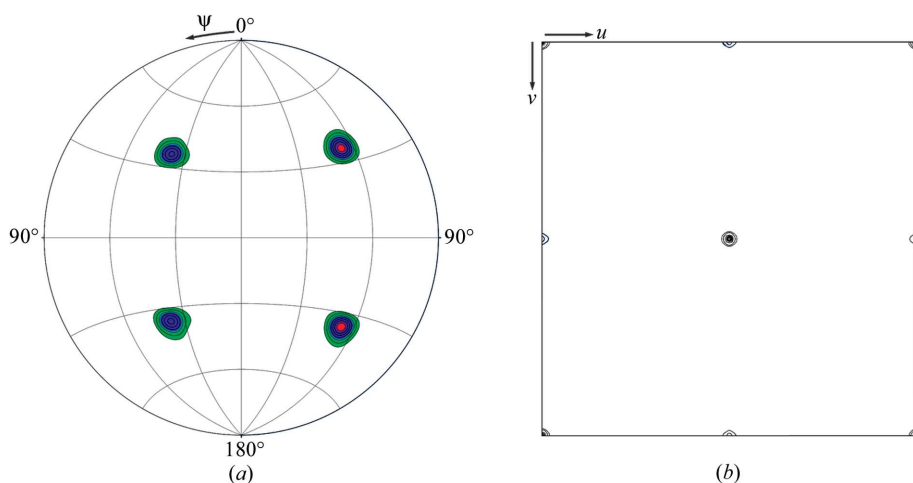
Expression, purification, crystallization and preliminary crystallographic analysis of recombinant full-length P4 protein from bacteriophage $\phi 13$ have been previously reported (Kainov *et al.*, 2003; Mancini *et al.*, 2003). To obtain the selenomethionyl protein, *Escherichia coli* strain BL21(DE3) was grown in LB media with $150 \mu\text{g ml}^{-1}$ ampicillin at 310 K to an A_{540} of 0.6, collected by centrifugation and resuspended in the same volume of MOPS minimal medium containing $50 \mu\text{g ml}^{-1}$ L-selenomethionine (Nanduri *et al.*, 2002). Cultures were induced by adding 1 mM IPTG and incubated for 14 h at 290 K. Protein purification was performed as for the native protein (Kainov *et al.*, 2003). Crystals of selenomethionyl $\phi 13$ P4 were grown at 293 K at a concentration of 12 mg ml^{-1} using 0.1 M Tris–HCl pH 7.0, 0.9 M trisodium citrate and 0.2 M NaCl as a precipitant and using the sitting-

drop vapour-diffusion method (Harlos, 1992). Crystals were cryoprotected by transferring them into a reservoir solution mixed with glycerol to a final concentration of 20% prior to freezing in a nitrogen-gas stream at 100 K. Using X-rays tuned to maximize the anomalous scattering signal from the Se atoms [$f'' = 5.4 \text{ e}^-$; as judged by fluorescence scan analysis (Evans & Pettifer, 2001) on the UK CRG beamline BM14 at the ESRF Grenoble (Mancini *et al.*, 2003); Table 1], a total of 208° of data were collected in a continuous sweep as a series of 0.5° oscillations. Data were acquired to a resolution of 2.5 \AA with a MAR 165 mm CCD detector using an exposure time of 10 s per frame. Despite the fact that BM14 is a bending-magnet beamline and far less intense than a typical insertion-device beamline on a third-generation synchrotron, it is notable that the entire experiment required only about 2 h. Data were processed and scaled using the *HKL2000* suite of programs (Otwinowski & Minor, 1997). Details of the data-collection statistics are summarized in Table 1; the data were collected with reasonable redundancy (3.1 million observations of 0.32 million unique reflections to 2.5 \AA resolution) and with good precision [$\langle I/\sigma(I) \rangle = 9.9$]. The data images could be integrated reasonably well in any of the triclinic, primitive and C-centred monoclinic and orthorhombic and primitive tetragonal lattices. However, the analysis of the reflection intensities during scaling indicated that the true space group was C2. Analysis of the moments of the normalized structure-factor amplitudes (E values) suggests that the crystal is not twinned; indeed, presumably owing to the special disposition of the molecules, the moments are slightly higher than expected for a random distribution of atoms. Thus, the second moment of E values for the acentric data is 2.3 (the expected value is 2 for untwinned and 1.5 for twinned data). The self-rotation function (Fig. 1a) indicates that the protein occurs as sixfold symmetric oligomers arranged in two distinct orientations.

In addition, the native Patterson function (Fig. 1b) shows a peak (at 15% of the origin height) suggesting a non-crystallographic translation between components of the crystallographic asymmetric unit. Together with the size of the protein (38 kDa per subunit) and the unit cell, these data suggest that four hexamers (*i.e.* 24 subunits) of P4 are present in each crystallographic asymmetric unit, with 54% solvent content.

4. Substructure determination using *SHELXD*

The P4 protein from bacteriophage $\phi 13$ contains six methionine residues in each subunit and the crystallographic asymmetric unit contains four hexamers. Consequently, in the selenomethionated protein 144 Se atoms are present in the crystallographic asym-

**Figure 1**

Self-rotation function and native Patterson. (a) $\kappa = 60^\circ$ section of the crystallographic self-rotation function. The clear peaks suggest a sixfold symmetric arrangement of protein chains in the crystal. (b) $w = 0$ section of the native Patterson map. In addition to the origin peak (and the crystallographically related peak at $u = 0.5, v = 0.5$), there are peaks at $u = 0.5, v = 0$ and $u = 0, v = 0.5$. Both maps were calculated using the *CNS* (Brünger *et al.*, 1998) and *CCP4* programs (Collaborative Computational Project, Number 4, 1994) and rendered using the software *GROBAT* (R. M. Esnouf, unpublished program).

Table 2Anomalous signal-to-noise ratios as a function of resolution, calculated using the program *XPREP* (Bruker AXS).

Resolution range (Å)	∞ –8.0	8.0–6.0	6.0–5.0	5.0–4.1	4.1–3.9	3.9–3.7	3.7–3.5	3.5–3.3	3.3–3.1	3.1–2.9	2.9–2.7	2.7–2.5
Anomalous signal-to-noise ratio	1.61	1.80	1.58	1.43	1.38	1.28	1.16	1.07	0.99	0.91	0.85	0.84

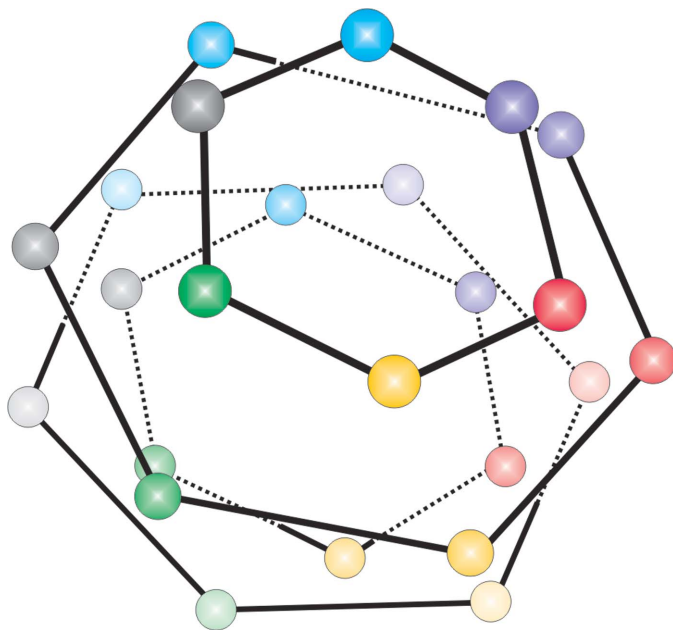
metric unit. The first task in phase determination is to determine the position of these atoms.

The computer program *XPREP* (Bruker AXS) was used to estimate pseudo- F_A values (where F_A are the structure factors of the anomalously scattering atoms) from the SAD data and indicated that there was a useful anomalous signal present in the data to 3.5 Å (*i.e.* anomalous signal > 1.2 times the noise; see Table 2). These F_A values were then fed into *SHELXD* using default parameter values for selenium-substructure determination (Sheldrick & Schneider, 1997), excluding data beyond 4.0 Å, which quickly yielded solutions consisting of four sixfold-symmetric sets, each with 24 atoms (see Fig. 2 for an example of such a hexameric structure), with CC/weakCC values of 48.2/27.2%. Given the size of the Se-atom substructure of the P4 protein in this crystal form, this appeared to be a remarkable result. To date few structures of such complexity have been solved from anomalous scattering data by direct methods (Deacon & Ealick, 1999; Weeks *et al.*, 2001). However, it became clear that these initial solutions were incorrect since subsequent phase refinement using the program *SHARP* (de La Fortelle & Bricogne, 1997) did not improve the phases and no interpretable electron-density map could be obtained after density modification (the final figure

of merit was 0.20 and the anomalous phasing power for the acentric reflections was 0.6).

However, by letting *SHELXD* run beyond the point at which, by general consensus, the correct solution had been found, further solutions were obtained that resulted in a very small increase in the CC from 48.22 to 48.36% (Fig. 3*a*). Although these solutions, like the earlier ones, located only 96 atoms, they behaved significantly better in phase refinement using *SHARP*, such that a total of 120 Se atoms (five in each of the 24 chains) could then be located in the crystallographic asymmetric unit (final figure of merit = 0.25 and anomalous phasing power = 0.9). These gave good starting phases for the calculation of an electron-density map into which, following 24-fold NCS averaging and solvent flattening (*GAP* program; JMG and DIS, unpublished work), the protein structure could be readily built (manuscript in preparation).

In several further calculations the same behaviour was observed: *SHELXD* repeatedly found several incorrect solutions before locking on to the correct one. All of these solutions contained four hexameric arrangements of Se atoms, but in the case of the incorrect solutions these hexamers were inverted, rotated or translated with respect to each other. Perhaps the most striking of these incorrect intermediate solutions was one in which the two pairs of hexamers were different enantiomorphs (Fig. 3*b*).

**Figure 2**

The hexameric arrangement of Se atoms found by dual-space direct phasing methods for the single-wavelength anomalous data set of bacteriophage ϕ 13 P4. Atoms are grouped and coloured by spatial proximity. Four layers of sixfold symmetric groups of atoms (connected by lines) are found along the sixfold symmetry axis of the arrangement, a total of 24 atoms.

5. Substructure determination using *SnB*

For direct comparison, the substructure determination of the same data set was also carried out using the program *SnB* using renormalized anomalous difference $|E|$ data and default parameters (Blessing & Smith, 1999; Weeks & Miller, 1999), giving remarkably similar results. As in the case of *SHELXD*, *SnB* found many solutions consisting of hexameric arrangements of Se atoms. Again, different types of solutions were obtained (see Fig. 4), only some of which correspond to the correct solution leading to interpretable electron-density maps. This is unlike most successful phase determinations reported so far, in which a single group of correct solutions can typically be identified (Weeks & Miller, 1999). As with *SHELXD*, correct and incorrect *SnB* solutions differ only in the relative spatial arrangement of the four Se-atom hexamers in the crystallographic asymmetric unit (see examples in Fig. 4). By running *SnB* for 1000 cycles, three clear peaks in the histogram of solutions sorted by R_{\min} were observed, two of which were well separated from the large peak of grossly incorrect solutions. The lowest R_{\min} (0.42) solutions correspond to the correct solution as found by *SHELXD* and the first correct solution came after 118 trials, compared with the related incorrect solutions ($R_{\min} = 0.43$), the first of which appeared at trial 62. Overall, therefore, the behaviour echoed

that of *SHELXD*: for this case, given sufficient trials the correct solution is found and can be detected on the basis of the standard quality indicators of the program; however, premature termination could lead to the selection of an incorrect solution.

6. An intricate false-minimum problem

Both *SnB* and *SHELXD* encounter similar problems in determining the substructure of the anomalous scatterers. Each produces many solutions which are partly correct (four hexamers of Se atoms are correctly represented, but there is an incorrect spatial relationship between and internal to these hexamers). In both cases, correct and incorrect solutions are very similar in their crystallographic quality indicators (*e.g.* Fig. 3*a*).

The work of *SHELXD* and *SnB* in reciprocal space is driven by an initial selection of the largest normalized *E* values. It is likely that these strong *E* values will be biased towards reflections from lattice planes of strong electron density

arising from layers of Se atoms within each oligomer (the non-crystallographic translations between hexamers will reinforce this effect). This will in turn bias the derived triplet invariants and thereby influence the phase refinement. *E* values that correspond to lattice planes in different orientations are likely to be smaller and have less influence on the phasing process. Since either enantiomorph is equally likely, the two pairs of hexamers that comprise the asymmetric unit can adopt quasi-independent hands. Thus, clusters of sets of atoms can be selected that internally largely satisfy the measures of correctness and yet may be mutually inconsistent.

It is also informative to consider the issue in Patterson space. For closely associated groups of atoms, such as the hexameric substructures of the P4 protein, vectors between atoms within each of the substructures will tend to pile up on each other since they occupy a relatively small volume of Patterson space close to the origin (the choice of enantiomorph for the substructure is immaterial since the Patterson function is centrosymmetric). These peaks therefore dominate the Patterson function and solutions will feature the correct

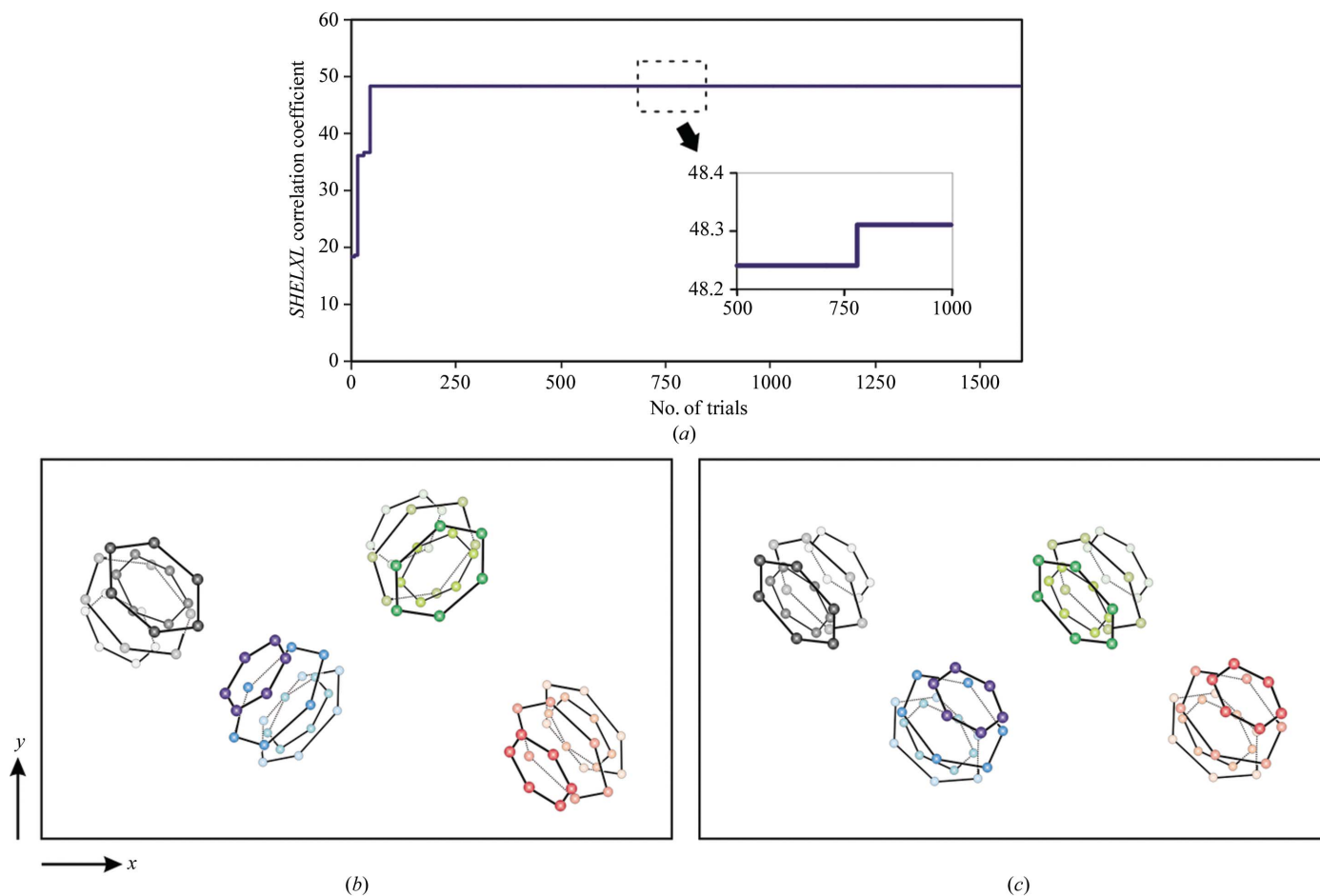


Figure 3

Examples of the heavy-atom substructure solutions of the bacteriophage $\phi 13$ P4 protein single-wavelength anomalous data set obtained from the program *SHELXD* (Sheldrick & Schneider, 1997). (*a*) The trajectory of correlation coefficients. After approximately 75 iterations, an initial solution [illustrated in (*b*); correlation coefficient 48.2%] is obtained. After ~ 800 cycles a solution with a slightly higher correlation coefficient is found [48.3%; solution shown in (*c*)]. (*b*) An early solution in which two out of four hexamers have inverted handedness (grey and red *versus* blue and green). (*c*) Correct solution obtained from a later cycle of *SHELXD*. All hexamers have the same handedness. Only the contents of one crystallographic asymmetric unit (four hexamers) are shown.

internal selenium substructures for the hexamers (irrespective of the choice of enantiomorph for each hexamer). The relative positions of these hexamers depend on vectors between atoms generally distant from one another in different hexamers in the asymmetric unit. These inter-hexamer vectors correspond to long Patterson vectors at lower density, close to the noise level, which therefore make a relatively small contribution. It is these long Patterson vectors that are important in locking the substructure of the oligomers into consistent sets. In the special case of $\phi 13$ P4 there is one point, far from the origin, where Patterson vectors pile up owing to the translation between pairs of hexamers. This will help ensure that the choices of enantiomorph for pairs of hexamers are locked together. Note that these translation peaks can *in extremis*

cause problems since a partial interpretation of the Patterson function can be obtained by treating them as arising from a single very heavy atom (the 'U-atom' effect).

We believe that these considerations explain why for $\phi 13$ P4 both *SnB* and *SHELXD* converge on false minima containing positional and orientational inversions and correct solutions in which all interatomic vectors match the peaks in the Patterson function. Furthermore, we can understand why the progression from a wrong solution to the correct one is not necessarily smooth, since unifying the enantiomorph of inconsistent partial substructures by the sequential correction of individual atoms will produce intermediate structures with some long-range vectors corrected but will introduce incorrect local vectors within the oligomer. It is therefore hardly surprising that even the powerful algorithms of *SHELXD* and *SnB* have difficulty in escaping from such local minima.

7. Conclusions

Direct phasing methods have a known tendency to converge on false minima, often owing to space-group-dependent factors. Here, we have described an example where false minima arise from the nature of the structure to be solved, namely a collection of oligomers arranged in a non-random fashion. Occurrences of the false-minima problem as described here for $\phi 13$ P4 are bound to increase as structural biologists tackle increasingly more complex biological systems that will inevitably contain many protein subunits, often related by a high degree of NCS. Implementing methods to circumvent false minima will be useful. Simple visual inspection of the false solutions can often immediately reveal their internal inconsistencies, suggesting that relatively simple algorithms may detect and perhaps alleviate the problem. One possible approach would be to detect clusters of atoms and to occasionally invert the enantiomorph of individual clusters of sites, perhaps strengthening the algorithm by incorporating

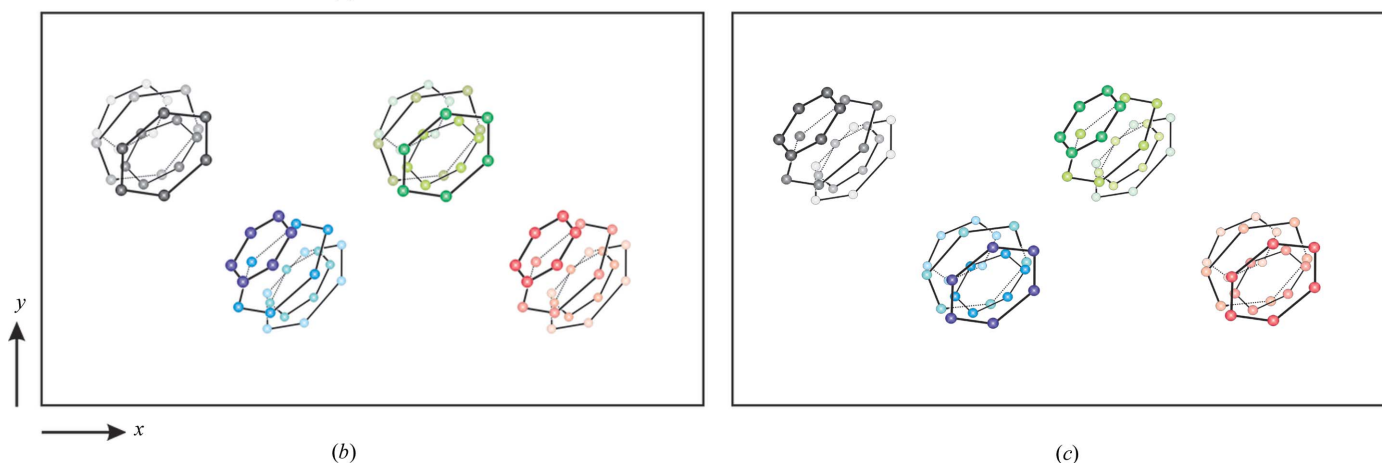
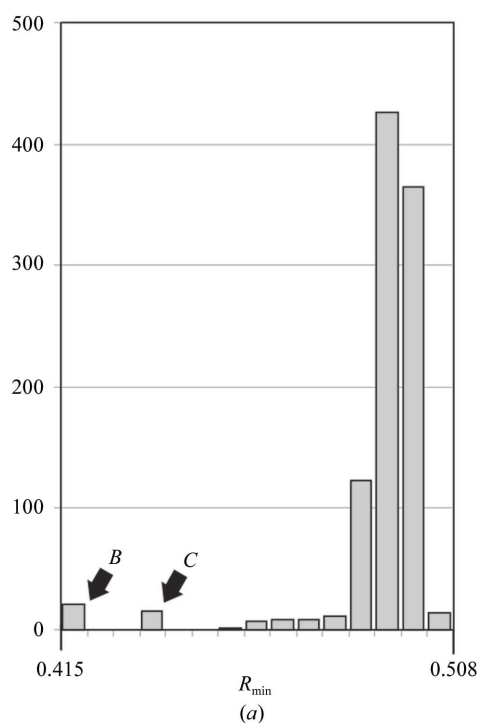


Figure 4

Histogram of R_{\min} values obtained from the program *SnB* (Weeks & Miller, 1999). (a) Apart from the large noise peak (highest R_{\min}), there are two significant peaks representing two classes of solutions. Analysis shows that only peak *B* (lowest R_{\min} values) contains solutions which have the correct representation of all four hexamers in the asymmetric unit (b). Peak *C* corresponds to false minima (c). 1000 iterations were performed in total. (b) A correct solution [from peak *B* in (a)]. (c) An incorrect solution [from peak *C* in (a)]: in this case the four hexamers have consistent enantiomorphs but are misplaced. Only the contents of one crystallographic asymmetric unit (four hexamers) are shown.

NCS-symmetry detection and filtering. Nevertheless, the example of $\phi 13$ P4 demonstrates the remarkable power of *Shake-and-Bake*-derived methods to deal with very complex substructures, provided that they are allowed to dig deeper than conventional wisdom might suggest.

We thank Geoff Sutton for key assistance at BM14. BM14 is supported by the UK Research Councils. Work was supported by the Human Frontiers Science Programme, the UK Medical Research Council and the Academy of Finland (Finnish Centre of Excellence Program 2000–2005, grants 1202855 and 1202108 to DHB). CM is funded by a Wellcome Trust studentship. EJM was supported by an EMBO postdoctoral fellowship (ALTF-192) and HFSP. JMG is supported by the Royal Society and DIS by the UK MRC.

References

- Blessing, R. H. & Smith, G. D. (1999). *J. Appl. Cryst.* **32**, 664–670.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* **D54**, 905–921.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Deacon, A., Weeks, C., Miller, R. & Ealick, S. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 9284–9289.
- Deacon, A. M. & Ealick, S. E. (1999). *Structure*, **7**, 161–166.
- Delft, F. von, Inoue, T., Saldanha, S. A., Ottenhof, H. H., Schmitzberger, F., Birch, L. M., Dhanaraj, V., Witty, M., Smith, A. G., Blundell, T. L. & Abell, C. (2003). *Structure*, **11**, 985–996.
- Evans, G. & Pettifer, R. F. (2001). *J. Appl. Cryst.* **34**, 82–86.
- Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 287–307.
- Harlos, K. (1992). *J. Appl. Cryst.* **25**, 536–538.
- Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
- Kainov, D. E., Pirttimaa, M., Tuma, R., Butcher, S. J., Thomas, G. J. Jr, Bamford, D. H. & Makeyev, E. V. (2003). *J. Biol. Chem.* **278**, 48084–48091.
- La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
- Mancini, E. J., Grimes, J. M., Malby, R., Sutton, G. C., Kainov, D. E., Juuti, J. T., Makeyev, E. V., Tuma, R., Bamford, D. H. & Stuart, D. I. (2003). *Acta Cryst.* **D59**, 2337–2341.
- Mancini, E. J., Kainov, D. E., Grimes, J. M., Tuma, R., Bamford, D. H. & Stuart, D. I. (2004). *Cell*, **118**, 743–755.
- Mancini, E. J., Kainov, D. E., Wei, H., Gottlieb, P., Tuma, R., Bamford, D. H., Stuart, D. I. & Grimes, J. M. (2004). *Acta Cryst.* **D60**, 588–590.
- Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.
- Mindich, L. (1999). *Microbiol. Mol. Biol. Rev.* **63**, 149–160.
- Nanduri, B., Byrd, A., Eoff, R. L., Tackett, A. J. & Raney, K. D. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 14722–14727.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* **D58**, 1772–1779.
- Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, R. & Uson, I. (2001). *International Tables for Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, pp 333–351. Dordrecht: Kluwer Academic Publishers.
- Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
- Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.
- Weeks, C. M., Sheldrick, G. M., Miller, R., Uson, I. & Hauptman, H. A. (2001). *Advances in Structure Analysis (XVIIIth European Crystallographic Meeting, 16–20 August 1998, Prague, Czech Republic)*, edited by R. Kusel & J. Hasek, pp 37–64. Prague: Czech & Slovak Crystallographic Association.
- Xu, H., Weeks, C. M., Deacon, A. M., Miller, R. & Hauptman, H. A. (2000). *Acta Cryst.* **A56**, 112–118.